

Data management protocols workshop

Discipline specific protocol: rate your reproducibility

Kees den Heijer

Data Steward faculty Civil Engineering and Geosciences, TU Delft

University of Tartu
April 10, 2019

L^AT_EX source revision 173 (April 10, 2019)



Protocols

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact

- What do you consider a protocol?
- Are protocols used in your field?

About us

Data Stewards TU Delft

Kees den Heijer

Data Steward
Faculty of Civil Engineering and Geosciences



Esther Plomp



Data Steward
Faculty of Applied Sciences

About the speaker

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

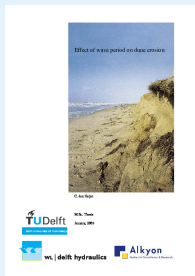
Contact

Experienced in data and information matter

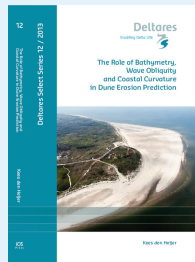
- Data Steward @ CEG faculty
- Owner / consultant @ Data2day



Educated in Coastal Engineering (TU Delft, CEG)



MSc (2005)



PhD (2013)

What is your data management like?

Introduction

Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

Contact



Outline

Discipline specific protocol: rate your reproducibility

- 1 Introduction
- 2 Open and reproducible research
- 3 Sand motor example
OpenEarth
Version Control
- 4 Onion model
- 5 Exercise
- 6 Contact

Open research data

Computational Fluid Dynamics (CFD) in drinking water treatment (Wols, 2010)

Introduction

Open and
reproducible
research

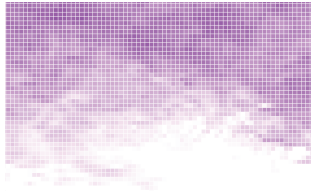
Sand motor
example

OpenEarth
Version Control

Onion model

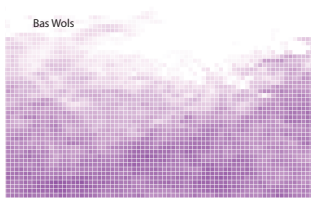
Exercise

Contact



CFD in drinking water treatment

Bas Wols



Open research data

Computational Fluid Dynamics (CFD) in drinking water treatment (Wols, 2010)

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact

List of publications

160

Data publications

Experimental and numerical data used in this thesis can be accessed online (through <http://dx.doi.org>).

All data:

<http://dx.doi.org/10.4121/uuid:c1ac7344-1419-4398-ba13-c757851c303f>

Movies of particle distributions in ozone systems (Figure 4.8):

<http://dx.doi.org/10.4121/uuid:f204af22-0442-431e-9896-6a04bdf1acfb>

Movies of spreading of dye around a single cross-flow UV lamp (Figure 5.3 & 5.4):

<http://dx.doi.org/10.4121/uuid:594f0e25-aa30-4546-a0fa-656f36bf72d1>

Movies of advanced oxidation around a single cross-flow UV lamp (Figure 5.18):

<http://dx.doi.org/10.4121/uuid:dff99405-75c4-455c-048a-d50b4e4a9087>

Movies of particle distributions in single cross-flow UV lamp systems (Figure 5.20):

<http://dx.doi.org/10.4121/uuid:62ef98c8-5a01-442a-9cac-80106576a171>

Movies of spreading of dye in a bench-scale UV reactor (Figure 6.4):

<http://dx.doi.org/10.4121/uuid:06096a3-3fa1-4502-825b-6527a5d6aa5f>

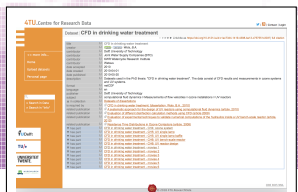
Movies of a LES model of a bench-scale UV reactor (Figure 7.1):

<http://dx.doi.org/10.4121/uuid:47499ab9-c354-4418-87c9-69b02acac43>



CFD in drinking water treatment

Bas Wols



Open and reproducible research

Aeolian Sediment Availability and Transport (Hoonhout, 2017)

Introduction

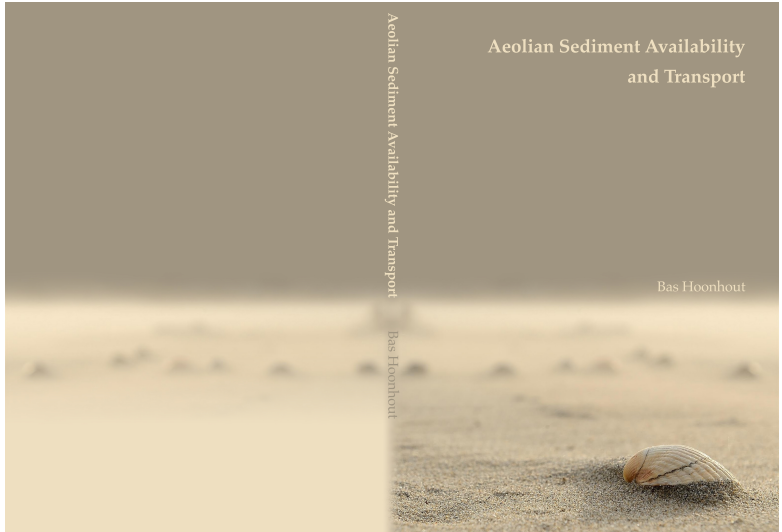
Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

Contact



Open and reproducible research

Aeolian Sediment Availability and Transport (Hoonhout, 2017)

Aeolian

Aeolian Sediment Availability and Transport

Bas Hoonhout

PhD

Ansible configuration to generate PhD thesis on aeolian sediment availability and transport (10.4233/uuid:e84894d6-87d2-4006-e8c2-d9fbfacabddc) from raw data.

Usage

```
>>> ansible-playbook playbook.yml
```

WARNING: This Ansible playbook takes about 2 weeks to run on an average computer

Sand Motor

Introduction

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact

Large scale experiment



photo 2011, by Joop van Houdt

Construction year 2011

Sand volume $21.5 \cdot 10^6 \text{ m}^3$

Research

- Various research grants
- 15 PhD researchers
- International cooperation
- Multi-disciplinary
 - Morphology
 - Biology
 - Ecology
 - Hydrology
 - Chemistry
 - Governance

Sand Motor

Introduction

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact

Large scale experiment



photo 2017, by Joop van Houdt

Construction year 2011

Sand volume $21.5 \cdot 10^6 \text{ m}^3$

Research

- Various research grants
- 15 PhD researchers
- International cooperation
- Multi-disciplinary
 - Morphology
 - Biology
 - Ecology
 - Hydrology
 - Chemistry
 - Governance

Types of data

Wide variety

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact



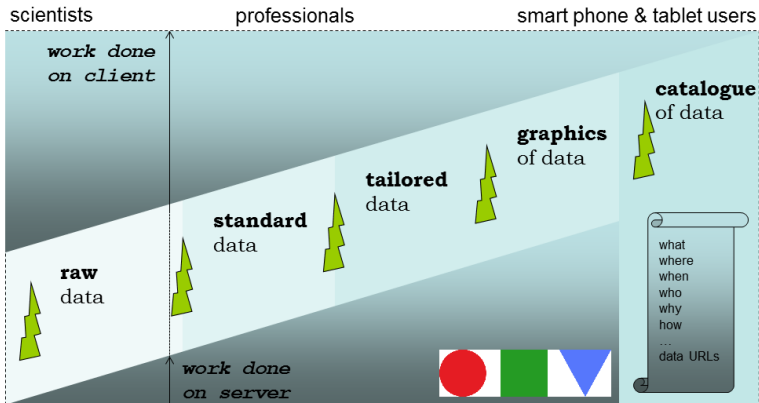
OpenEarth

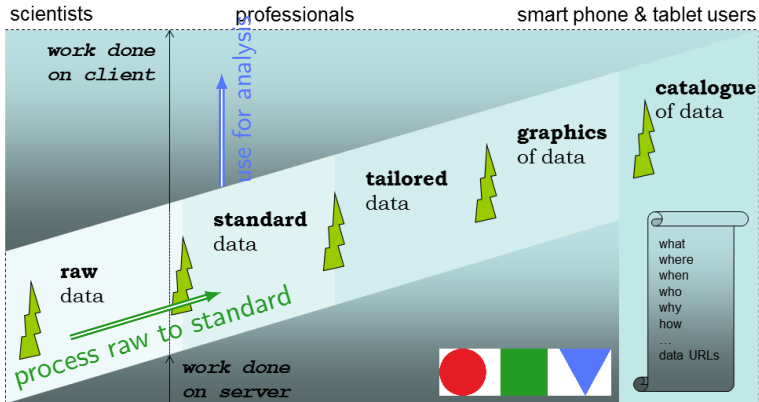
Community

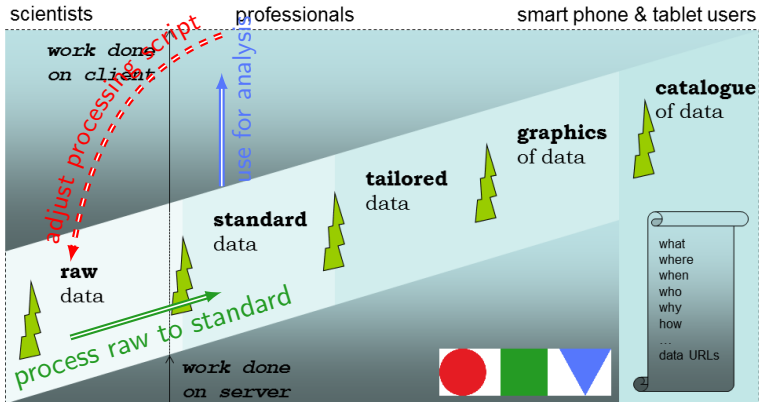
- LinkedIn group (729 members)
- [http://www.openearth.eu / wiki](http://www.openearth.eu/wiki)
- <https://github.com/openearth>
- subversion repositories (<https://svn.oss.deltares.nl/>):
 - OpenEarth Tools
 - OpenEarth Models (schematizations)
 - OpenEarth Raw Data

Highlights

- Open source software
- Open sharing of data, models and tools
- Reproducible workflows
- Version control (svn/git) for raw data, models and tools
- Open standards (OGC, CF, netCDF)







Graphics

Sand motor topography/bathymetry

Introduction

Open and
reproducible
research

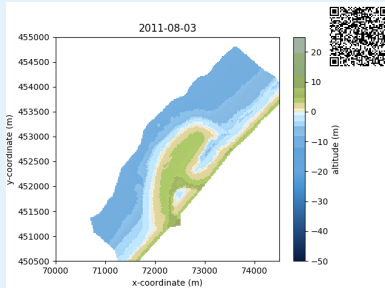
Sand motor
example
OpenEarth
Version Control

Onion model

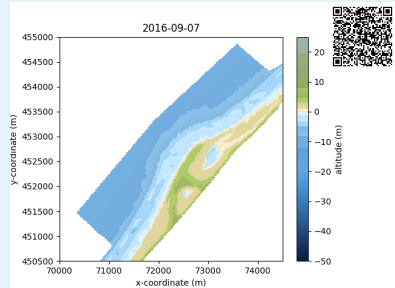
Exercise

Contact

2011



2016



Graphics

How created, with reference to source data

Introduction

Open and
reproducible
researchSand motor
exampleOpenEarth
Version Control

Onion model

Exercise

Contact

zandmotor_bathy.py Revision 14141	zandmotor_bathy.py Revision 14142
<pre> 1 # -*- coding: utf-8 -*- 2 """ 3 4 \$Id: zandmotor_bathy.py 14141 2018-02-02 14:12:01Z heijer \$ 5 \$Date: 2018-02-02 15:12:01 +0100 (Fri, 02 Feb 2018) \$ 6 \$Author: heijer \$ 7 \$Revision: 14141 \$ 8 \$HeadURL: https://svn.oss.deltares.nl/repos/openeearthtools/trunk/pytho >h/OpenEarthTools/openeearthtools/plot/zandmotor_bathy.py \$ 9 10 \$author: heijer 11 """ 12 13 import netCDF4 14 import numpy as np 15 import matplotlib.pyplot as plt 16 import matplotlib.gridspec as gridspec 17 import qrcode 18 19 from colormap_vaklodigen import vaklodigen_colormap 20 21 # https://doi.org/10.4121/uuid:c40da555-3eff-4c3c-89d6-136994a07120 22 nc_url = 'http://opendap.tudelft.nl/thredds/dodsC/data2/zandmotor/morp >hology/JETSKI/gridded/jetskikb18_3736.nc' 23 idx = 0 24 25 26 def label(netcdf_var): 27 longname = netcdf_var.long_name 28 units = netcdf_var.units 29 return '%s (%s)' % (longname, units) 30 31 32 def add_qrcode2fig(fig, size=.18): 33 keywords = {'\$HeadURL': 'https://svn.oss.deltares.nl/repos/openeearth >ools/trunk/python/OpenEarthTools/openeearthtools/plot/zandmotor_bathy. >py \$', '\$Revision: 14141 \$', } 34 for i, keyword in enumerate(keywords): 35 keywords[i] = keyword.replace('\$', '').replace('Revision: ', ' >').replace('HeadURL: ', '').strip() 36 img = qrcode.make('%s?p=%s' % (keywords[0], keywords[1])) </pre>	<pre> 1 # -*- coding: utf-8 -*- 2 """ 3 4 \$Id: zandmotor_bathy.py 14142 2018-02-02 14:13:05Z heijer \$ 5 \$Date: 2018-02-02 15:13:05 +0100 (Fri, 02 Feb 2018) \$ 6 \$Author: heijer \$ 7 \$Revision: 14142 \$ 8 \$HeadURL: https://svn.oss.deltares.nl/repos/openeearthtools/trunk/pytho >h/OpenEarthTools/openeearthtools/plot/zandmotor_bathy.py \$ 9 10 \$author: heijer 11 """ 12 13 import netCDF4 14 import numpy as np 15 import matplotlib.pyplot as plt 16 import matplotlib.gridspec as gridspec 17 import qrcode 18 19 from colormap_vaklodigen import vaklodigen_colormap 20 21 # https://doi.org/10.4121/uuid:c40da555-3eff-4c3c-89d6-136994a07120 22 nc_url = 'http://opendap.tudelft.nl/thredds/dodsC/data2/zandmotor/morp >hology/JETSKI/gridded/jetskikb18_3736.nc' 23 idx = -1 24 25 26 def label(netcdf_var): 27 longname = netcdf_var.long_name 28 units = netcdf_var.units 29 return '%s (%s)' % (longname, units) 30 31 32 def add_qrcode2fig(fig, size=.18): 33 keywords = {'\$HeadURL': 'https://svn.oss.deltares.nl/repos/openeearth >ools/trunk/python/OpenEarthTools/openeearthtools/plot/zandmotor_bathy. >py \$', '\$Revision: 14142 \$', } 34 for i, keyword in enumerate(keywords): 35 keywords[i] = keyword.replace('\$', '').replace('Revision: ', ' >').replace('HeadURL: ', '').strip() 36 img = qrcode.make('%s?p=%s' % (keywords[0], keywords[1])) </pre>

Tailored data

Published with DOI

Introduction

Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

Contact

4TU Centre for Research Data

Dataset | Sand Motor gridded bathymetric and topographic survey

<< more info...

Home

Upload datasets

Personal page

Search in Data

Search in "info"

TU Delft

TU/e

UNIVERSITEIT TWENTE

WAGENINGEN

title

creator

creator

creator

contributor

date accepted

date created

date published

description

subject

time coverage

in collection

calculated from

description

subject

time coverage

spatial coverage

calculated from

description

subject

spatial coverage

spatial coverage

map

related publication

related publication

licence

Sand Motor gridded bathymetric and topographic survey

de Zeeuw, R.

de Zeeuw, R.

de Vries, S. (Sierd)

den Heijer, C. (Kees)

TU Delft, Faculty of Civil Engineering and Geosciences, Department of Hydraulic Engineering

2017-01-20

2016-12-06T15:29Z

2017

Zandmotor bathymetric and topographic survey, gridded on 20m grid

Delfland Echo sounder ERC advanced grant 291206-NEMO GPS interpolation on grid Jetski NatureCoast Sand Motor Shore

years 2011-2016 2011-08-03T00:00Z/2016-09-07T00:00Z

Zandmotor data

Nelmo morphology data survey path Delfland 2012-2016

Morphological data collected within the TU Delft Nelmo project. Bimonthly measurements of marine and aeolian morphology of the cross shore coastal profile over an alongshore extent of 17km. Note that the Zandmotor Jetski dataset (DOI!!!!!!) is complementary to this dataset. This is non interpolated data (measured survey paths).

Nelmo survey path

years 2012-2016 2012-02-25T00:00:00Z/2016-09-07T00:00:00Z

Coverage of Nelmo morphology data survey path Delfland 2012-2016, Zandmotor

Sand Motor Topographic Survey, actual surveyed path

Bathymetry and topography at the Zandmotor after construction

survey path

Zandmotor Topographic Survey, surveyed path

Coverage of Zandmotor gridded bathymetric and topographic survey

Map [kml]

https://doi.org/10.1016/j.coastaleng.2015.10.011

https://doi.org/10.1016/j.coastaleng.2016.03.005

https://doi.org/10.2112/JCOASTRES-D-13-00070.1

General terms of use

Direct access through OPeNDAP

Total 250 MiB

[Move over links for options and info]

jetskib117_3938.nc

jetskib117_3938.nc

jetskib117_3938.nc

top of page

Access on our OPeNDAP server: xml, html

© 2018 4TU.ResearchData

ORE RDF/XML

Tailored data

How created, with reference to source data

Introduction

Open and
reproducible
research

Sand motor
example

OpenEarth
Version Control

Onion model

Exercise

Contact

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jul 29 21:38:30 2014

$Id$
$Date$
$Author$
$Revision$
$HeadURL$

@author: heijer
"""
import datetime
id = '$Id$'
idsplit = id.split(' ')
frmt = '%Y-%m-%dT%H:%M:%SZ'
history = 'netCDF created on: ' + str(datetime.datetime.utcnow().strftime(frmt)) + ' with script ' + idsplit[1] + ', script revision number is ' + idsplit[4] + ' .'

from python.jetski_ncgen import NCgen
import os

# initialize class
nc = NCgen()

# adjust settings
nc.schema.attributes['comment'] = ''
nc.schema.attributes['history'] = history
nc.schema.attributes['license'] = 'DISCLAIMER: This data is made available in the hope that it will be useful, but WITHOUT ANY WARRANTY.
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.'
# Creator Search
nc.schema.attributes['creator_name'] = 'R. de Zeeuw, M. de Schipper, S. de Vries'
nc.schema.attributes['creator_url'] = 'https://www.shoremonitoring.nl and https://www.tudelft.nl'
nc.schema.attributes['creator_email'] = 'roeland@shoremonitoring.nl, m.a.deschipper@tudelft.nl, siord.devries@tudelft.nl'
nc.schema.attributes['institution'] = 'Shore Monitoring & Research and Technische Universiteit Delft'
nc.schema.attributes['project'] = 'Zandmotor monitoring'
# Publisher Search
nc.schema.attributes['publisher_name'] = 'Kees den Heijer'
nc.schema.attributes['publisher_email'] = 'c.denheijer@tudelft.nl'
nc.schema.attributes['publisher_url'] = 'https://www.tudelft.nl'

# grid settings
nc.schema.grid.tiletype = 'vaklodingen'
nc.schema.grid.tileprefix = 'jetskiKB'
# path settings
nc.schema.ncpath = os.path.join(os.path.dirname(__file__), '..', '..', 'results')
if not os.path.exists(nc.schema.ncpath):
    os.mkdir(nc.schema.ncpath)
nc.raw.path = os.path.join(os.path.dirname(__file__), '..', 'raw')

# run data processing
nc.process_data()
```

What is your data management like?

Introduction

Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

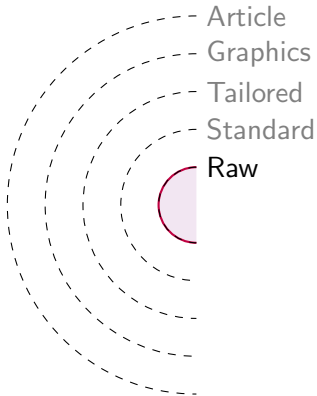
Onion model

Exercise

Contact



Raw data



Just as collected at the sensor.
Essential information for
interpretation should be attached:

- sensor type
- location
- circumstances
- logbook

Standard data

Introduction

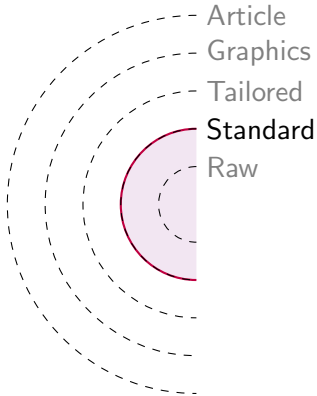
Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

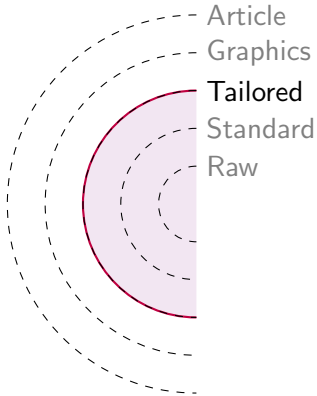
Contact



Standardised in terms of:

- format
- units
- lineage
- metadata (vocabularies)

Tailored data



Tailored to particular research questions and/or (re)use

- Aggregation
- Analysis
- Combination

Graphics

Introduction

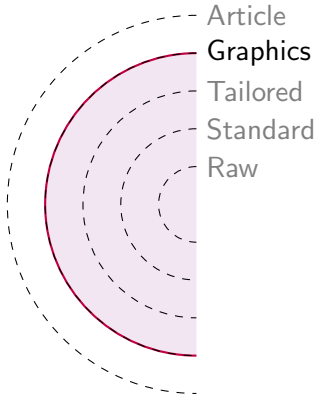
Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

Contact



Visualisation of standard and/or tailored data.
Optionally interactive.

Conclusions / Article

Introduction

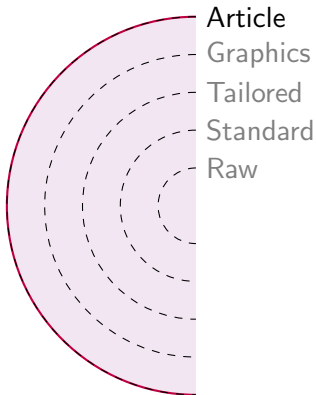
Open and
reproducible
research

Sand motor
example
OpenEarth
Version Control

Onion model

Exercise

Contact



Wrap-up of the research, referring to all other layers and explaining how they interrelate and provide evidence for the conclusions.

Sketch your data workflow

Include for instance:

- observation data
- processing
- interpretation

Exercise

Sketch your data workflow

Include for instance:

- observation data
- processing
- interpretation

Prioritise

Indicate key items in your workflow, based on e.g.:

- data volume
- possibility to recreate/redo
- time/resources needed to recreate/redo

Questions?

name Kees den Heijer

room 2.36 or at a CEG department

availability Tuesday - Friday

phone +31 15 27 85607

email C.denHeijer@tudelft.nl / DataSteward-CiTG@tudelft.nl

ORCID  <https://orcid.org/0000-0003-0314-2779>

LinkedIn  <https://www.linkedin.com/in/keesdenheijer/>